Engineering
Simplicity

Network Configuration Example

IP Fabric Upgrade Minimum Operating Procedure

Published

2023-05-11

Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, California 94089
USA
408-745-2000
www.juniper.net

Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

Network Configuration Example IP Fabric Upgrade Minimum Operating Procedure

The information in this document is current as of the date on the title page.

YEAR 2000 NOTICE

Juniper Networks hardware and software products are Year 2000 compliant. Junos OS has no known time-related limitations through the year 2038. However, the NTP application is known to have some difficulty in the year 2036.

END USER LICENSE AGREEMENT

Table of Contents

CHAPTER 1 IP Fabric Upgrade Overview

About This Configuration Example

Overview

Use this network configuration example (NCE) to upgrade all the switches in an IP Fabric architecture that is already up and running.

Documentation Feedback

We encourage you to provide feedback so that we can improve our documentation.
Send your comments to design-center-comments@juniper.net. Include the document or topic name, URL or page number, and software version (if applicable).

CHAPTER 2 Plan Upgrade for Switches in an IP Fabric Architecture

Planning the Upgrade

This section covers the guidelines for planning the upgrade.

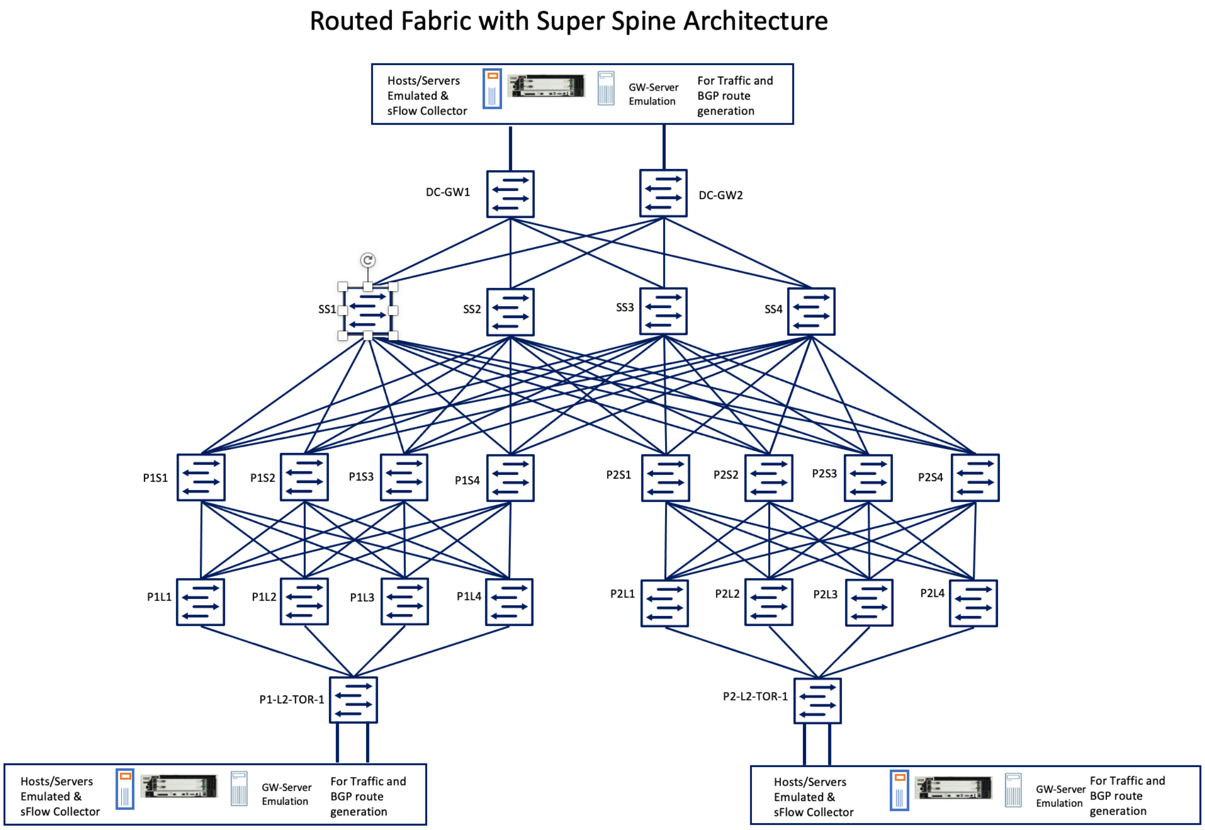
- Always upgrade one device at a time at initial phase. After a few successful upgrades and developing familiarity with the procedure, you can upgrade switches in batches, with multiple leaf switches at a time, for large deployment. It is recommended that spine and super-spine switches must be upgraded one at a time, to prevent any disruption to traffic in case there are no redundant paths.
- Check the current bandwidth usage of all network links.
- Both in-band and out-of-band upgrade procedures can be used. If even a single switch in the IP fabric uses ZTP based upgrades via in-band procedure, DHCP relay needs to be configured on all switches in the IP Fabric. This is to ensure that the switch being upgraded has continued access to DHCP server for software image and switch configuration download. If all the switches in the IP Fabric are being upgraded through ISSU/NSSU, there is no need to configure DHCP relay on any switch in IP Fabric for in-band procedure.
- There are some CLI show commands that are used, and information is collected. It is recommended to automate the CLI commands in scripts and store all the information on the server. Use tools to quickly search and compare details with the collected information.
- Identify and design some traffic flows that can validate the network connectivity and can be running in background during the upgrade. Ping and traceroute are commonly used for this purpose.
- Plan enough time for the maintenance window (MW). Multiple MWs might be needed for large deployment. Always upgrade one device at a time at initial phase. After a few successful upgrades and getting familiar with the procedure, upgrade can be done in batches with multiple leaf switches at a time for large deployment.
- Schedule the change with all teams and individuals who are affected by the change.
- This document supports multi-homing of a server host to different top-of-rack (TORs) or leaf switches using LACP based LAG connection.

CHAPTER 3 Deployment Architectures

DC IP Routed Fabric 5 Stage Clos with Super Spine

This architecture includes DC IP Routed Fabric 5 Stage Clos with Super Spine with eBGP as fabric protocol with every layer in different AS.

Figure 1: IP Fabric Topology with EBGP Fabric Protocol, with Every Layer in Different AS



Routed Fabric with Super Spine Architecture

Hosts/Servers Emulated & sFlow Collector
GW-Server Emulation
For Traffic and BGP route generation
DC-GW1
DC-GW2
SS1
SS2
SS3
SS4
P1S1
P1S2
P1S3
P1S4
P2S1
P2S2
P2S3
P2S4
P1L1
P1L2
P1L3
P1L4
P2L1
P2L2
P2L3
P2L4
P1-L2-TOR-1
P2-L2-TOR-1
Hosts/Servers Emulated & sFlow Collector
GW-Server Emulation
For Traffic and BGP route generation
Hosts/Servers Emulated & sFlow Collector
GW-Server Emulation
For Traffic and BGP route generation

Supported Platforms

Table 1 lists the supported platforms for different roles in IP Fabric.

Table 1: Supported Platforms for IP Fabric

Device Roles	Platforms
Leaf/TOR	• QFX5130-32CD • QFX5220-32CD /128C • QFX5120-32C/48Y/48T/48YM • QFX5100/QFX5110-48S • QFX5200-32C • QFX5210-64C • ACX7100-48L
Spine	• QFX5220-32CD/128C • PTX10K8 /16

Device Roles	Platforms
	• QFX5120-32C • QFX5210-64C • QFX5130-32CD • QFX5700 • QFX5200-32C • PTX10003 • QFX5110-32Q
Super Spine	• QFX5220-128C • PTX10K8/PTX10K3 • QFX5210-64C • QFX10K8/16

NOTE: In the listed supported platforms, PTX10K8/16, QFX5700, QFX10K8/16 are chassis based modular systems. Remaining platforms are fixed form factor of size 1, 2, or 3 Rack units (RUs).

Node Roles

In Figure 1, the following are the node roles:

- P1L1, P1L2, P1L3 and P1L4 are the leaf nodes in POD-1.
- P2L2, P2L2, P2L3 and P2L4 are the leaf nodes in POD-2.
- P1S1, P1S2, P1S3, P1S4 are the spine nodes in POD-1.
- P2S1, P2S2, P2S3, P2S4 are the spine nodes in POD-2.
- SS1, SS2, SS3, SS4 are the super spines in super spine layer, that is common to both POD-1 and POD-2.

Switch Configurations

It is expected that the IP Fabric is running the basic minimum configurations that are required to make it functional. See Figure 1. Here is the minimal description of the configuration requirements:

- The fabric is configured for dual stack with both IPv4 and IPv6 routing in the fabric.
- All the links in the fabric are P2P and configured for both IPv4 and IPv6.
- eBGP is being used as the routing protocol.
- eBGP peer groups are used, all leaf switches belong to 1 eBGP peer group (say, LEAF), all spine switches belong to 1 eBGP peer group (say, SPINE) and all the super-spine switches belong to 1 eBGP peer group (say SUPER-SPINE).
- The routes might be aggregated before being advertised through eBGP.
- Bidirectional traffic is flowing through all switches in the IP fabric.

CHAPTER 4 Manual Upgrade for Switches Without Juniper Apstra

Guidelines for Layer by Layer Upgrade

At each layer, identify all commonly used platforms, so that any upgrade related exceptions for other platforms can be identified.

First Step - Upgrade TORs (Edge Switches)

Upgrade all TORs one by one. It is assumed that all TORs are single RE devices and therefore the TOR being upgraded is unavailable during the upgrade for data path forwarding. In case a server is connected to only one TOR which is being upgraded, migrate VMs onto servers that are connected to TORs that are not being upgraded.

Second Step - Upgrade Leaf Devices

Upgrade all leaf switches, one by one: leaf1, leaf2, leaf3, and so on. For dual RE switches, use the ISSU or NSSU procedure.

Third Step - Upgrade Spines

Upgrade all spine switches, one by one: spine1, spine2, spine3, and so on. For dual RE switches, use the ISSU or NSSU procedure.

Fourth Step - Upgrade Super-Spines

Upgrade all super-spine switches, one by one: super-spine1, super-spine2, super-spine3, and so on. For dual RE switches, use the ISSU or NSSU procedure.

Common Procedure for Switch Upgrade

Pre-Upgrade and Post-Upgrade Health Checking

We recommend the health checking of the switch being upgraded, both pre-upgrade and post-upgrade, and recording this information. Recorded pre-upgrade and post-upgrade health check information can be compared in the event of an issue.

Health Check Procedure

Perform the following steps:

1. Check whether user traffic flow is as expected, without any loss before and after the upgrade.
2. Backup all devices' configurations and save them on a server before the upgrade.
3. Collect detailed information and check the system is in healthy state before and after the upgrade.
    - Check syslog for any failure and errors

      ```
      show log messages | no-more
      show log chassisd | no-more
      ```

- Check alarms and core-dump on the system

  ```
  show chassis alarms | no-more
  show system alarms | no-more
  show system core-dumps | no-more
  ```

- Check RE/FPC/PIC status and interfaces status (for all platforms that are supported)

  ```
  show chassis hardware detail | no-more
  show chassis fpc detail | no-more
  show chassis fpc pic-status | no-more
  show chassis environment | no-more
  show chassis routing-engine | no-more
  show interfaces descriptions | match up | no-more
  show interfaces descriptions | match down | no-more
  show interface xe-* | match "physical|rate" | no-more
  show interface et-* | match "physical|rate" | no-more
  ```

  On chassis based modular platforms, you can also run the fabric related CLIs:

  ```
  show chassis fabric summary | no-more
  show chassis fabric fpcs | no-more
  ```

  For dual RE switches, we need to check switchover readiness for ISSU/NSSU:

  a. In case of dual RE switches, the backup RE should be GRES ready.

  b. Check Master RE : "`Switchover Ready`" ready status by running the command "`request chassis routing-engine master switch check`"

  c. Run "`show system switchover`" command in Backup RE and ensure it is in ready state.

- Check routing table and forwarding table, these should be as expected:

  ```
  show system processes extensive | no-more
  show krt queue | no-more
  show route summary | no-more
  show route forwarding-table summary | no-more
  show arp no-resolve expiration-time | no-more
  ```

  In the ARP CLI above, no-resolve indicates that we don't want to perform DNS lookups for every entry in the ARP table. So, with no-resolve, we only see IP addresses, which can be faster when you have several ARP entries.

- Check and store detailed information of IP fabric related information (for all supported platforms):

  - `show pfe statistics traffic | no-more`
  - `show system virtual-memory | no-more`
  - `show task memory detail | no-more`
  - `show system memory | no-more`
  - `show task memory | no-more`
  - `show chassis fpc | no-more`
  - `show chassis routing-engine | no-more`
  - `show system processes extensive | no-more`
  - `show system processes memory detail | no-more`
  - `show bgp summary | no-more`
  - `show interfaces ae* terse | no-more`

- show lacp interfaces
- show lacp statistics interfaces
- show interfaces terse |no-more
- show bfd session | no-more

On chassis-based modular platforms, you can run Switch Interface Board (SIB) related command:

```
show chassis sibs | no-more
```

Perform any customized checks and preparation procedures for the planned upgrade device, after finishing the health check. These checks must be done before the upgrade procedure listed in the following sections of this document.

Preparation for the Upgrade

Perform the following steps:

1. Check system free storage to ensure enough storage for the new Junos OS image:
   - Run “*df -k /var/tmp*” at shell mode for checking free space.
   - In case, free space is less than required space for upgrade, then we can free up space using command listed in step 2. For ZTP, there is no such free space requirement.
2. Run `"request system storage cleanup dry-run"` next to check the proposed list of files for deletion:
   - Use “*request system storage cleanup* “ command to free storage space on the devices if the proposed list of files to be deleted is acceptable.
3. Clear any alarms and core-dumps before the upgrade:
   - clear system errors fpc all fpc-slot <fpc slot id>
4. Copy Junos OS image to device `/var/tmp` directory. Use this step only if phone-home or ZTP is not used.

Pre-Upgrade BGP Specific Operations on Spine

Perform the following steps:

1. It is assumed that each switch being upgraded has following BGP parameters preconfigured:
   - delay-route-advertisements minimum-delay inbound-convergence
   - delay-route-advertisements minimum-delay routing-uptime

   This is to ensure that the BGP routes are advertised by a switch after power on, only after route convergence has been assured at the local switch. This means that routes in RIB have been downloaded to FIB of the switch. In case the routes in RIB are unavailable at FIB, the local router starts advertising routes immediately after routes are available in RIB, even though the FIB cannot forward them. It might lead to traffic drop for destinations whose routes were advertised by local router, but for which there is no corresponding route in FIB of the local router.

The definition of the BGP parameters is as follows:

a) inbound convergence – Specify a minimum delay in route advertisement after the source peer has sent all route updates to the local router being upgraded. The local device being upgraded waits for at least the configured duration after inbound convergence at local device has completed for the source peer. For BGP routes, the source peer sends the end-of-rib after all the route updates have been sent to local device. The default value is 120 seconds, the range is 1 through 36000 seconds.

   If all the BGP peers of the device are of type IPv4, run the following command:

   -
     ```
     set protocols bgp family inet unicast delay-route-advertisements minimum-delay inbound-convergence <1 to 36000 s>
     ```

     If all the BGP peers of the device are of type IPv6, run the following command:

   -
     ```
     set protocols bgp family inet6 unicast delay-route-advertisements minimum-delay inbound-convergence <1 to 36000 s>
     ```

     If some of the BGP peers of the device are of type IPv4, and some are of type IPv6, then inbound-convergence on local device needs to be set on a per-peer basis. For IPv4 BGP peers, run the following command:

   -
     ```
     set protocols bgp group <group_name> neighbor <neighbor IPv4 address> family inet unicast delay-route-advertisements minimum-delay inbound-convergence <1 to 36000 s>
     ```

     For IPv6 BGP peers, run the following command:

   -
     ```
     set protocols bgp group <group_name> neighbor <neighbor IPv6 address> family inet6 unicast delay-route-advertisements minimum-delay inbound-convergence <1 to 36000 s>
     ```

b) minimum routing uptime - Specify the minimum delay, in seconds, before sending a route advertisement after the routing protocol process (rpd) starts. The device waits for at least the configured duration before sending out route advertisements to its peers. The default value is 0 seconds, the range is 1 through 36000 seconds.

   If all the BGP peers of the device are of type IPv4, run the following command:

   -
     ```
     set protocols bgp family inet unicast delay-route-advertisements minimum-delay routing-uptime <1 to 36000 s>
     ```

If all the BGP peers of the device are of type IPv6, run the following command:

```
o   set protocols bgp family inet6 unicast delay-route-
    advertisements minimum-delay routing-uptime <1 to 36000 s>
```

If some of the BGP peers of the device are of type IPv4, and some are of type IPv6, then `routing-uptime` on local device must be set on a per-peer basis. For IPv4 BGP peers, run the following command:

```
o   set protocols bgp group <group_name> neighbor <neighbor
    IPv4 address> family inet unicast delay-route-
    advertisements minimum-delay routing-uptime <1 to 36000 s>
```

For IPv6 BGP peers, run the following command:

```
o   set protocols bgp group <group_name> neighbor <neighbor
    IPv6 address> family inet6 unicast delay-route-
    advertisements minimum-delay routing-uptime <1 to 36000 s>
```

For more information, see, https://www.juniper.net/documentation/us/en/software/junos/bgp/topics/ref/statement/delay-route-advertisements-edit-protocols-group-family-unicast.html

2. If any BGP peer switch sends traffic to the device, note the incremental egress traffic statistics (output packets) on its switch's device connected interfaces. In case there is an aggregated ethernet interface, between this peer switch and device, identify the constituent physical interfaces using the following command on peer switch:

```
o   show lacp interfaces <ae interface on peer switch connected to
    DUT>
```

Monitor the egress traffic rate on physical interfaces of peer switch, connected to device, using the following commands:

```
o   show interfaces <interface name of peer switch connected to
    DUT> | grep rate
o   monitor interface traffic
```

3. Create a policy for route rejection:

```
o   Set policy-options policy-statement <policy_name = DENY-ALL>
    then reject
```

4. If device is configured with BGP, withdraw all advertised BGP routes from all peer super-spines. Here, `SUPER-SPINES` group refers to all super-spine switches acting as BGP peers of spine switch being upgraded:

```
o   set protocols bgp group SUPER-SPINES export DENY-ALL and
    commit.
```

For information on the BGP route export commands, see https://www.juniper.net/documentation/us/en/software/junos/bgp/topics/ref/statement/export-edit-protocols-bgp.html

5. If device is configured with BGP, withdraw all advertised BGP routes from all peer leaves, here LEAF group refers to leaf switches acting as BGP peers of spine switch being upgraded:

```
o   set protocols bgp group LEAF export DENY-ALL and commit.
```

6. Note the IP address on device that is configured on its each BGP peer switch connected interface. Run the following command on device:

```
o   Show interfaces <interface to peer switch> brief
```

7. Verify on each BGP peer switch (both super-spine and leaf) that the routes being exported by device are withdrawn:

```
o   Show bgp summary
```

There are multiple entries displayed. Check the entry corresponding to device. This can be identified using the IP address configured on the device interface connected to the peer switch, on which this CLI was run. In the entry corresponding to device (on peer switch), the routes received from device should be shown as `0/0/0/0` under the column `State|#Active/Received/Accepted/Damped`.

Alternatively, you can also run the following command on each BGP peer switch (both super-spine and leaf) of device:

```
o   Show bgp neigbor <DUT interface address facing BGP peer switch>
```

This should show the same information under the heading: Table inet.0

8. If any BGP peer switch sends traffic to the device, monitor traffic statistics on that peer switch and wait until the incremental egress stats (output packets) on its device connected interfaces become almost 0. In case there is an aggregated ethernet interface between this peer switch and device, identify the constituent interfaces using the following command on peer switch:

```
o   show lacp interfaces <ae interface on peer switch connected to
    DUT>
```

Monitor the egress traffic rate on device connected interfaces of peer switch using the following command:

```
o   show interfaces <interface name of peer switch connected to
    DUT> | grep rate
o   monitor interface traffic
```

Pre-Upgrade BGP Specific Operations on Leaf

Perform the following steps:

1. Follow step 1 of pre-upgrade BGP specific operation on spine switch.

2. The procedure for withdrawing advertised BGP routes on leaf switch being upgraded, is very similar to that of spine switch. Withdraw advertised routes to all spines acting as BGP peers:

```
o   set protocols bgp group SPINES export DENY-ALL and commit.
```

Here, `SPINES` refers to BGP peer group that is configured on leaf switch for peering with all the spines via BGP.

3. In case a TOR switch (or server host) is connected through L2 MC-LAG to device leaf switch, disable the physical interface on leaf switch connected to that TOR switch (or server host). This is because the TOR switch (or server host) might not have eBGP configured and work based on L2 hashing of north-bound traffic to all leaf switches. Therefore, disable the interface on device leaf switch being upgraded, towards TOR switch

(or server host). This prevents TOR switch (or server host) from sending any traffic to device leaf switch being upgraded.

   - `set interfaces <interface name> disable and commit.`

4. Follow the steps 5 through 7 in case of pre-upgrade BGP operations on spines.

Pre-Upgrade BGP Specific Operations on Super-Spines

Perform the following steps:

1. Follow step 1 of pre-upgrade BGP specific operation on spine switch.
2. The procedure for withdrawing advertised BGP routes on super-spine switch being upgraded, is similar to that of leaf switch. Withdraw advertised routes to all spines acting as BGP peers:

   - `set protocols bgp group SPINES export DENY-ALL and commit.`

   Here, `SPINES` refers to BGP peer group that is configured on super-spine switch for peering with all the spines through BGP. Follow steps 5 through 7 in case of pre-upgrade BGP operations on spines.

Upgrade Device with New Image and Reboot

There are various methods for upgrading a switch with new image:

- CLI based upgrade
- ZTP
- ISSU
- NSSU

The detailed descriptions are available in the section, Manual Upgrade Details for Switches Without Juniper Apstra.

Post-Upgrade Routines Common to Single RE and Dual RE Switches And All Switch Roles

Perform the following steps:

1. Wait and verify the device is up.
2. Verify that there are no process cores.

   - `show system core-dumps`

3. Verify that there are no additional system and chassis alarms.

   - `show chassis alarms | no-more`
   - `show system alarms | no-more`

Post-Upgrade BGP Specific Operations on Spine

Perform the following steps:

1. Restart advertising routes to all peer spines. Here, `SUPER-SPINES` group refers to super-spine switches acting as BGP peers of spine switch being upgraded:
   - o `delete protocols bgp group SUPER-SPINES export DENY-ALL and commit.`
2. Restart advertising routes to all peer leaves. Here, `LEAF` group refers to leaf switches acting as BGP peers of spine switch being upgraded:
   - o `delete protocols bgp group LEAF export DENY-ALL and commit.`
3. In case traffic was sent from any spine (acting as BGP peer switch) to the device, monitor traffic statistics on that peer switch and wait until the incremental egress statistics (output packets) on its device connected interfaces become almost pre-upgrade value. In case there is an aggregated ethernet interface, between this peer switch and device, identify the constituent physical interfaces using this CLI on peer switch:
   - o `show lacp interfaces <ae interface on peer switch connected to DUT>`

   Monitor the egress traffic rate on physical interfaces of peer switch connected to spine switch being upgraded using the following CLIs:

   - o `show interfaces <interface name of peer switch connected to DUT> | grep rate`
   - o `monitor interface traffic`

Post-Upgrade BGP Specific Operations on Leaf and ToR Switch / Server Host

Perform the following steps:

1. Restart advertising BGP routes on leaf switch being upgraded. This step is similar to that of spine switch. Restart advertising routes to all spines acting as BGP peers:
   - o `delete protocols bgp group SPINES export DENY-ALL and commit.`

   Here, SPINES refers to BGP peer group that is configured on leaf switch for peering with all the spines via BGP.
2. In case a TOR switch (or server host) is connected to leaf switches via L2 MC-LAG, the physical interface on leaf switch connected to the TOR switch (or server host), needs to be re-enabled (if it was disabled earlier). This re-enables TOR switch (or server host) to send the traffic to all the leaf switches (including leaf switch being upgraded) after L2-hashing.
   - o `set interfaces <interface name> enable and commit.`
3. Follow step 3 in case of post-upgrade BGP operations on spines.

Post-Upgrade BGP Specific Operations on Super-Spines

Perform the following steps:

1. Restart advertising BGP routes on super-spine switch being upgraded. This step is similar to that of spine switch. Restart advertising routes to all spines acting as BGP peers:

   - `delete protocols bgp group SPINES export DENY-ALL and commit.`

   Here, `SPINES` refers to BGP peer group that is configured on super-spine switch for peering with all the spines via BGP. Follow step 3 in case of post-upgrade BGP operations on spines.

2. Follow step 3 in case of post-upgrade BGP operations on spines.

Verify All Network Core-Facing Interfaces

Perform the following steps:

1. Wait and verify all underlay routing are up.
2. Wait and verify BGP neighbour relationships are established. Wait for BGP route update to finish.

   a) `"show bgp summary"` and check Established status for all neighbours.

3. Wait and verify IRB interfaces are up. This is applicable only if IRB interfaces are configured, this typically happens on ToR or CE switches.

   a) `show interfaces irb`

Verify all End-Device Facing Access Interfaces

Wait and verify all user traffic is resumed as normal.

Post-Upgrade Health Check

Repeat the health check procedure that was performed before the upgrade. It can be followed by any customized checks.

Post-Upgrade Cleanup

1. Delete the newly installed image if required.
2. Restore syslog configuration setting back to original.

Supported Platforms for Manual Upgrade Procedures (Without Juniper Apstra)

Table 2 provides the details of the supported platforms.

Table 2 Manual Upgrade Procedure

Upgrade Method	Supported Platform Reference
ISSU	https://apps.juniper.net/feature-explorer/issu.html
NSSU	https://apps.juniper.net/feature-explorer/feature-info.html?fKey=1175&fn=Nonstop+software+upgrade+%28NSSU%29
ZTP	https://apps.juniper.net/feature-explorer/parent-feature-info.html?pFKey=1272&pFName=Zero+Touch+Provisioning

Manual Upgrade Details for Switches Without Juniper Apstra

Both Single and Dual RE Switches

Normal CLI based upgrade can be used for both single and dual RE switches. This is the simplest upgrade option and lacks the features supported by other options. First, ftp the new software package to `/var/tmp` directory on device. Next, run the following command:

```
root@host> request system software add </var/tmp/new_package_name>
reboot
```

Single RE Switches Only

Zero Touch Provisioning (ZTP) restores a switch to factory default configuration. In ZTP, pre-existing configuration must be reapplied through a file server where the Junos OS Evolved or Junos OS image is stored, as it will be lost. Note that the switch being upgraded after ZTP must be connected to the configuration server / image server under all circumstances, so that pre-existing configuration can be reapplied after ZTP is over.

Assumptions

The device uses information that is configured on a Dynamic Host Configuration Protocol (DHCP) server to locate the necessary software image and configuration files on the network. If the DHCP server is not configured to provide this information, then the preinstalled software and default factory configuration are loaded.

This document assumes that dhcpd, vsftpd, tftpd, and httpd are installed and configured to support ZTP. The device that is provisioned for downloading image and configuration files uses one of vsftpd, httpd, and tftpd. DHCP is used to provide options for ZTP.

DHCP Relay

If the ZTP server and the device to be upgraded are not directly connected on the same LAN, a DHCP relay is required. The relay must be enabled on any device that provides connectivity between the device being upgraded and the ZTP server using following CLIs:

```
Set forwarding-options dhcp-relay server-group test <IP address of
DHCP server>
Set forwarding-options dhcp-relay active-server-group test
Set forwarding-options dhcp-relay group all interface <list of
interfaces on which DHCP discover messages are expected>
```

For more information on configuring a DHCP relay on a Junos OS device, see the following documents:

- https://www.juniper.net/documentation/us/en/software/junos/dhcp/topics/topic-map/dhcp-relay-agent-security-devices.html
- https://www.juniper.net/documentation/us/en/software/junos/dhcp/topics/topic-map/dhcpv6-relay-agent.html
- https://www.juniper.net/documentation/en_US/junos/topics/topic-map/dhcpv6-relay-agent-switching-devices.html

Setting Up DHCP Server and Transport Mode

Perform the following steps:

1. Refer to https://linux.die.net/man/5/dhcpd.conf to know more about parameters and declarations. Below is a sample config of `/etc/dhcp/dhcpd.conf`.

   - # interface over which dhcp server listens to dhcp discover messages.

```
      DHCPDARGS=ens33;

# The below declaration is used to identify the subnets over which
      to listen
# for dhcp discover messages and provide ip addresses.
# range : specifies how many ip addresses to lease.
# domain-name : is used to identify the network, domain name-
      servers: used
# when host names are used instead of ip addresses.

      subnet 3.3.3.0 netmask 255.255.255.0 {
          range 3.3.3.3 3.3.3.15;
          option domain-name “mydomain.net";
          option domain-name-servers 10.209.194.133;
          option routers 3.3.3.254;
          default-lease-time 60000;
          max-lease-time 720000;
      }
# Below declaration provides an option space definition.

      option space SUNW;
      option SUNW.server-image code 0 = text;
      option SUNW.server-file  code 1 = text;
```

```
       option SUNW.image-file-type code 2 = text;
       option SUNW.transfer-mode code 3 = text;
       option SUNW.symlink-server-image code 4 = text;
       option SUNW.http-port  code 5 = text;
       option SUNW-encapsulation code 43 = encapsulate SUNW;

       # group is used to apply common parameters for a bunch of
       different hosts.
       # defining a particular host and its parameters.
       # “hardware ethernet <mac>” mac-address of the device. For
       MX10003 it will
       # have the mac address of the fxp0 interface.
       # “transfer-mode <mode>” mode used for downloading the
       image and config
# files. If this absent, default is tftp. Options are http, ftp and
       tftp.
       # log-server and ntp-server are for sending syslog
       messages.
       # “server-image <imagename>”  is the image for the device.
       # “server-file <filename>” is the option for the config
       file.
       # “tftp-server-name” is the ip address of the server that
       provides the files
       # for booting. This is provided as a string.

       group {
           next-server 3.3.3.1;
           host mx204-12345 {
       hardware ethernet 98:a4:04:7f:1a:83;
       option SUNW.transfer-mode "ftp";
       option host-name “mx204-12345";
       option log-servers 3.3.3.1;
       option ntp-servers 66.129.255.62;
       option SUNW.server-file   “dut-baseline-config.conf”;
       option SUNW.server-image  “junos-vmhost-install-mx-x86-64-
19.4R1.1.tgz”;
       option tftp-server-name "3.3.3.1";
           }
       }
```

Adhere to the text or number format as mentioned above. If not, dhcpd indicates an error upon startup. Save the config file and start the dhcpd service. The logs pertaining to dhcpd can be viewed in the `/var/log/messages` file.

2. Copy the image and configuration file to the appropriate paths depending on transport mode configured. The table below is an example assuming `/tftpboot/` is used by tftp and ftp for file store. The `server-file` and `server-image` options in `dhcpd.conf` file need to have the path relative to the path configured for the transport mode.

Transport Mode	Config File Path	Home Directory
ftp	/etc/vsftpd/vsftpd.conf	/tftpboot
tftp	/etc/xinet.d/tftp	/tftpboot
http	/etc/http/conf/httpd.conf	/var/www/html/

   For example, if the image is in `/tftpboot/PLATFORM_AA/image_aa.tgz`, then the server-image option must be `/PLATFORM_AA/image_aa.tgz`.

3. If a factory default device is being provisioned, make the network connections and power on the device. When the device boots, auto image upgrade (AIU) starts.

4. If an existing device is to be provisioned, it is best to zeroize the device using the "`request system zeroize`" command. Type "`yes`" for the prompt and press Enter.

   The device is zeroized and then rebooted. The device comes up in amnesiac mode. Login using root and there is no password prompt. After a couple of minutes, there are messages on the console indicating that ZTP has started. Issue "`show dhcp client binding`" CLI command to verify DHCP bound IP.

Monitoring ZTP Progress

Perform the following steps:

1. The following messages indicate options that are sent by the DHCP server:

```
Auto Image Upgrade: DHCP INET Options for client interface fxp0.0
       ConfigFile:
baseline_mt-bona ImageFile: junos-vmhost-install-mx-x86-64-
       20.3R1.3.tgz
       Gateway: 17.17.34.1 DHCP Server: 17.17.34.1 File Server:
       17.17.34.1
```

   Then, AIU uses information in the DHCP options to download the image and configuration files. The image is then installed. After image install step, AIU configures an event-option to apply the configuration from the downloaded configuration file. As the last step after the new image is installed, apply the configuration.

   Below is a snapshot of the messages that are displayed on the console after DHCP options are received.

```
Auto Image Upgrade: To stop, on CLI apply
"delete chassis auto-image-upgrade"  and commit
Auto Image Upgrade: Active on INET client interface : fxp0.0
Auto Image Upgrade: Interface::   "fxp0"
Auto Image Upgrade: Server::      "17.17.34.1"
Auto Image Upgrade: Image File::  "junos-vmhost-install-mx-x86-64-
20.3R1.3.tgz"
Auto Image Upgrade: Config File:: "baseline_mt-bona"
Auto Image Upgrade: Gateway::     "17.17.34.1"
Auto Image Upgrade: Protocol::    "ftp"

Auto Image Upgrade: FTP timeout set to 300 seconds
```

The following are the messages shown when image is downloaded and installed:

```
Auto Image Upgrade: Start fetching baseline_mt-bona file from server
17.17.34.1 through fxp0 using ftp
Auto Image Upgrade: File baseline_mt-bona fetched from server
17.17.34.1 through fxp0
Auto Image Upgrade: FTP timeout set to 300 seconds
Auto Image Upgrade: Start fetching junos-vmhost-install-mx-x86-64-
20.3R1.3.tgz file from server 17.17.34.1 through fxp0 using ftp
Auto Image Upgrade: File junos-vmhost-install-mx-x86-64-20.3R1.3.tgz
fetched from server 17.17.34.1 through fxp0
Auto Image Upgrade: Aborting image installation of junos-vmhost-
install-mx-x86-64-20.3R1.3.tgz received from 17.17.34.1 through
fxp0: Installed and fetched image version same
Auto Image Upgrade: Applying baseline_mt-bona file configuration
fetched from server 17.17.34.1 through fxp0
```

The following are messages from the `/var/log/messages` file that show the IP address being allocated to the device.

```
Sep 26 04:11:41 mx-phs-server1 dhcpd: DHCPREQUEST for 17.17.34.110
from e4:fc:82:0f:d2:00 (TC3718210039) via eth1
Sep 26 04:11:42 mx-phs-server1 dhcpd: DHCPACK on 17.17.34.110 to
e4:fc:82:0f:d2:00 (TC3718210039) via eth1
Sep 26 05:11:41 mx-phs-server1 dhcpd: Vendor-Class-Identifier:
Juniper:ex4600-40f:TC3718210039
Sep 26 05:11:42 mx-phs-server1 dhcpd: DHCPREQUEST for 17.17.34.110
from e4:fc:82:0f:d2:00 (TC3718210039) via eth1
Sep 26 05:11:42 mx-phs-server1 dhcpd: DHCPACK on 17.17.34.110 to
e4:fc:82:0f:d2:00 (TC3718210039) via eth1
```

As the last step after the image is installed, apply the configuration. Run the `“show system commit”` to verify the output. The device must include a valid configuration at the end. For detailed installation logs, you can check the `/var/log/image_load_log` file.

Verification

To verify the device has received IP address from DHCP server, run the following command:

```
root@host> show dhcp client binding
```

In the output, the DHCP state must display “BOUND”.

```
root@host>show log image_load_log
```

The output shows the progress of ZTP image loading process.

Troubleshooting

Perform the following steps:

1. Ensure that the connections are working. Since the DHCP discover messages are broadcast, the network must forward these DHCP discover messages to the DHCP server.
2. The dhcpd process status must be running or active. If not, check the `/var/log/messages` file to verify the issue. Use the same file to look for DHCP entries to verify if the DHCP discover messages reach the DHCP server. At this point, the device should be assigned an IP address.

3. The DHCP messages in `/var/log/messages` should pertain to the mac-address of the `fxp0/em0` interface. If it is not present, then the DHCP discover messages from the device are not reaching the server.

4. Verify that the `fxp0/em0` interface receives IP address as shown in the command output `“show dhcp client binding”.`

   In addition to the IP address, the device must receive information about the image file, configuration file, server IP, and the transport mode to be used to provision the device.

   If only IP address is received without options, ensure that “tftp-server-name” option or the “server-name” options are present. If either of these two are not present, dhcpd does not send the additional options. If changes are made to any of the configuration files, the corresponding service must be restarted for the changes to take effect.

5. If the options are received but there are issues with downloading the image or configuration file, check the configuration for the corresponding service. The sample configurations and settings are shown below for a Centos 6.x installation.

Sample vsftd.conf options that are enabled for supporting ztp.

```
anonymous_enable=YES
local_enable=YES
local_root=/tftpboot/
write_enable=YES
local_umask=022
anon_upload_enable=YES
anon_mkdir_write_enable=YES
dirmessage_enable=YES
xferlog_enable=YES
xferlog_std_format=YES
ascii_upload_enable=YES
ascii_download_enable=YES

allow_writeable_chroot=YES
ls_recurse_enable=YES
listen=YES

pam_service_name=vsftpd
userlist_enable=NO
userlist_deny=NO
tcp_wrappers=YES
anon_root=/tftpboot/
```

Sample httpd.conf options for supporting ztp

```
ServerRoot "/etc/httpd"
Listen <ipaddress>:<port>
User daemon
Group daemon
EnableSendfile on
```

Sample tftp options for supporting ztp in /etc/xinetd.d/tftp

```
server_args  = -s /tftpboot/
disable      = no
```

6. More recent Linux distributions (for example, Centos 7 or later) have `firewalld` running by default. Configure to allow access for these services.

   Further details of this procedure can be found at https://www.juniper.net/documentation/us/en/software/junos/junos-install-upgrade/topics/topic-map/zero-touch-provision.html

Dual RE Switches Only

The ZTP procedure can be used for dual RE switches. There are other procedures available for dual RE switches as well.

NSSU

NSSU is for Virtual chassis (VC) and Virtual Chassis Fabric (VCF). They might play the role of a leaf switch in IP Fabric. It is expected that servers or CE switches are dual-homed to multiple line-cards (FPCs) of the leaf switches. So, the upgrade of each FPC in a VC/VCF does not disrupt the traffic or cause any service outage. GRES must happen as part of NSSU. For NSSU information, see: https://www.juniper.net/documentation/us/en/software/junos/high-availability/topics/concept/nssu-ex-series.html

For the NSSU procedure for VC/VCF formed using QFX5100 and EX switches, see: https://www.juniper.net/documentation/us/en/software/junos/high-availability/topics/topic-map/nssu-performing.html

For NSSU, we support upgrade from version N to N+3. For example, if current Junos OS version is 19, then the recommended upgrade version is maximum of 19+3 = 22.

ISSU

ISSU is used only for dual RE switches and combines GRES with NSR. The assumptions are as follows:

- Disk space is available for the /var file system on both Routing Engines
- Configuration is supported by a unified ISSU
- PICs are supported by a unified ISSU
- Graceful Routing Engine switchover is enabled
- Nonstop active routing is enabled

For VC/VCF, ISSU is not available.

For the ISSU information, see https://www.juniper.net/documentation/us/en/software/junos/high-availability/topics/topic-map/issu-understanding.html.

For the requirements to perform ISSU, see https://www.juniper.net/documentation/us/en/software/junos/high-availability/topics/concept/issu-system-requirements.html

For the detailed instructions to perform ISSU, see https://www.juniper.net/documentation/us/en/software/junos/high-availability/topics/topic-map/issu-performing.html

For ISSU, we support upgrade from version N to N+3. For example, if current Junos OS version is 19, then the recommended upgrade version is maximum of 19+3 = 22.

> **NOTE:** Some of the legacy PICs do not support ISSU. The PICs that go offline after ISSU need to be turned online manually.

CHAPTER 5 Juniper Apstra Based Upgrade

Upgrade Switch with Juniper Apstra Software

The procedure for draining a traffic from a switch via Apstra is specified here: https://www.juniper.net/documentation/us/en/software/apstra4.1/apstra-user-guide/topics/task/device-drain.html.

The video for the same is available at: https://www.youtube.com/watch?v=cpk-0eZ_L_U.

For a switch upgrade procedure, see https://www.juniper.net/documentation/us/en/software/apstra4.1/apstra-user-guide/topics/topic-map/device-nos-upgrade.html.

CHAPTER 6 Junos OS Software Rollback

Rollback Junos OS Software

Perform the following steps:

1. Contact customer support for alarms and core-dumps during/after the upgrade.
   - Provide the syslog file "messages" and core files
2. In case the switch upgrade failed due to ISSU, the switch automatically rolls back to its original Junos OS / Junos OS Evolved image. In case of NSSU, it is possible that some switches belonging to VC/VCF are successfully upgraded to newer Junos OS image, while the remaining switches did not. In such cases, you must manually rollback the newly upgraded switches to the original Junos OS image through the following commands:
   - `request system software rollback`
   - `request system reboot`

   Then, you can contact customer support for help in upgrading the switches that failed the upgrade process.